群智感知网络个性化位置隐私保护算法 *

胡　敏，张　艳，黄宏程†

(重庆邮电大学 通信与信息工程学院, 重庆 400065)

摘　要：群智感知网络中现有隐私保护算法对所有位置采用相同的隐私保护策略，导致位置隐私或保护过度或保护不足，且获得的感知数据精度较低。针对这一问题，提出了一种满足用户个性化隐私安全需求的位置隐私保护算法。首先，根据用户的历史移动轨迹，挖掘用户对不同位置的访问时长、访问频率以及访问的规律性来预测位置对用户的社会属性；然后，结合位置的自然属性，预测用户—位置的敏感等级；最后，结合用户在不同的位置有不同的隐私安全需求的特点，设置动态的隐私判定方案，在每个位置选敏感度低的用户参与感知任务，以确保用户在隐私安全的前提下，贡献时空相关性精确高的感知数据。仿真结果表明，该算法在提高隐私保护水平的同时还提高了感知数据的精度。

关键词：位置隐私保护；个性化；敏感等级；群智感知

中图分类号：TP309.2　　doi: 10.3969/j.issn.1001-3695.2017.10.0941

Personalized location privacy protection algorithm in crowd sensing networks

Hu Min, Zhang Yan, Huang Hongcheng†

(*School of Communication & Information Engineering*, *Chongqing University of Posts & Telecommunications*, *Chongqing* 400065, *China*)

Abstract: The existing privacy protection strategies in crowd sensing networks used the same privacy policies for all locations which overprotected led to the problems that some locations, others were not adequately protected and the sensing data was less accurate. In order to solve this problem, this paper proposed a location privacy protection algorithm to meet the users' personalized privacy and security requirements. First, it mined users' access duration, frequency and regularity at different locations according to the user's historical movement trajectory, which used to predict the social attributes of the locations to the users. Then, it combined the location's social attributes and natural attributes to predict user-location sensitivity levels. Finally, considering the different privacy security requirements of users in different locations, it set a dynamic privacy decision scheme. Users with less sensitivity at each location were selected to participate in sensing tasks to ensure that users, in the safe privacy context, could contribute the accurate data with a higher level of spatiotemporal correlation. The simulation results show that the algorithm can improve the privacy protection level and the accuracy of the sensing data.

Key Words: location privacy protection; personalization; sensitive level; crowd sensing networks

0　引言

群智感知是指通过人们已有的移动设备形成交互式的、参与式的感知网络，并将感知任务发布给网络中的个体或群体来完成，从而帮助专业人员或公众收集数据、分析信息和共享知识[1]。利用现有的感知设备和已有的通信网络构建群智感知网络，无须额外部署和维护感知设备以及传输网络，因此，群智感知网络能以低成本实现大规模和细粒度的感知，广泛应用于智能交通、社交网络、环境监测和健康监测等领域[2]。

由于大部分的感知数据都需要含有相关的时空位置等信息才有价值，而这些信息既直接包含了用户的隐私信息，又隐含了用户想要保护的其他敏感信息，如家庭住址、生活习惯、健康状况和社会关系等，从而造成用户隐私信息的泄露[3]。群智感知网络在为人们带来巨大便利的同时也带来了前所未有的隐私安全隐患，如果用户的隐私不能得到很好的保护，用户就很可能不再愿意共享他们的感知数据[4]。因此研究群智感知网络中的隐私保护，尤其是研究参与时空位置相关的感知任务的用户的位置隐私保护，具有重要的意义。

1　相关工作

至今已有不少文献研究了群智感知网络中位置隐私的保护方法，位置匿名就是常用的方法。Mehta 等人[5]利用虚假位置或

基金项目：重庆市科委基础与前沿研究项目（cstc2014jcyjA40039）

作者简介：胡敏（1971-），女，重庆人，副教授，硕士，主要研究方向为通信网体系与协议、无线通信等；张艳（1992-），女，湖北十堰人，硕士研究生，主要研究方向为群智感知网络；黄宏程（1979-），男（通信作者），河南南阳人，副教授，博士，主要研究方向为社会感知与社会计算(huanghc@cqupt.edu.cn)．

真实位置周围相关的语义位置来替代真实位置，向服务器提交位置数据。该方法有效保护了位置隐私，但是严重影响了感知数据的服务质量，降低了用户体验。文献[6]采用 *K*-匿名方法来保护位置隐私，即用一个包含 K 个用户的空间区域替代用户的真实位置，在这 K 个用户中，任何一个用户的位置都与其他 K-1 个用户的位置不可分辨。但是这种方法引入了严重的空间失真，不满足对感知数据效用的要求[7]。Yu 等人[8]提出了 *l*-多样性的原则，使满足 K-匿名的每个等价类中的敏感属性至少有 *l* 个值。*l*-多样性避免了一个等价类中敏感属性取值单一的情况，使隐私泄露风险不超过 1/*l*，但容易受到相似性攻击[9]。

文献[10]采用差分隐私的保护方法，往数据中添加随机噪声，保证攻击者能获取的个人数据几乎与它们没有这个人记录的数据集中能获取的相差无几，从而达到保护隐私的目的。但由于它们使用的原始拉普拉斯噪声是无界的，这使得发布无意义，造成用户隐私信息的泄露，妨碍数据发布的效用。为了解决这个问题，文献[11]提出了一种具有有界拉普拉噪声生成算法的差异化数据发布算法。这种噪声生成机制使得添加到真实位置数据中的噪声在合适范围内被采样以实现差分隐私。该方法大大减少了隐私的损失量，并提高了数据发布的效用。

另一种常用的位置隐私保护方法是通过加密技术保护用户的位置隐私。Wei 等人[12]提出了一种基于基站的属性基加密算法，结合假名技术实现了用户身份与数据存储的隔离。文献[13]提出一种基于同态加密技术的安全框架，确保用户的位置信息不会被发布给任何人，但是系统仍然可以将任务分配给位于每个感知任务位置附近的感知用户。该方法能有效保护用户的位置隐私，但是由于群智感知网络中的节点具有强移动性，加/解密时的密钥更新开销过大，密钥分发的过程比较复杂。

文献[14]提出的静态隐私保护机制采用位置混淆和数据隐藏的技术，假定所有位置所需要的隐私保护级别相同，因此，满足平均隐私阈值 θ 的平均静态（avg static）算法和满足最大隐私阈值 θ 的最大静态（max static）算法，均使用静态固定的隐私保护参数。这种使用静态固定参数的隐私保护技术虽然为用户提供了一些保护，却是以牺牲数据价值为代价的，没有考虑数据的效用。为了解决这个问题，文献[15]提出了一种自适应的位置隐私保护方法（Adaptive），在静态隐私保护策略的基础上，增加了感知应用的效用，综合考虑效用与隐私之间的权衡。但是这种方法依然忽略了不同用户在不同地点的个性化隐私需求问题。

不同的用户，其所关注的隐私信息内容可能不一样；即使是同一个感知任务，同一个用户在不同的位置对自身数据的隐私安全需求也不一样。若用同样的标准去提供相同级别的隐私保护，势必会导致用户在有些位置的隐私保护不足，造成隐私泄露，而另一些位置的隐私保护过度，造成感知资源的浪费。此外，对时空相关性要求较高的感知场景往往需要参与者提交精度较高的感知数据，而现有的位置隐私保护方法为了保护用户的隐私，提供粗粒度的感知数据，会大大降低服务质量。为了解决这个问题，本文提出了一种满足不同用户个性化隐私安全需求的位置隐私保护算法（location privacy protection algorithm for personalized security requirements of different users, LPPA-PSRDU），根据用户的历史轨迹信息，通过实时的敏感度计算，为感知用户提供一种动态的隐私判定方案，在每个位置选敏感度低的用户参与感知，同时提供时空位置精确的感知数据，在保护参与者位置隐私的同时，提高感知数据的服务质量。

2 个性化的隐私保护算法

本文提出的针对用户个性化隐私安全需求的隐私保护算法，主要考虑到不同的参与者对同一位置的敏感程度不同，通过用户的历史轨迹记录实时地评判出用户对该位置的敏感等级，尽量让敏感度低的用户来完成该位置的数据采集，以达到满足所有用户个性化的隐私安全需求的目的，同时提供高精度的感知数据。

为了合理地评判出某个感知用户对某个位置的敏感程度，考虑该位置所具有的自然属性和该位置对于用户来说所具有的社会属性两个方面。

定义 1 自然属性。自然属性指某一位置对于所有公共用户所具有的统一功能状态。例如，医院这个位置对所有病人来说都有相同的固有自然属性，不会因人而异。

定义 2 社会属性。社会属性指某一位置对于某个用户来说具有的特定属性，会因人的社会关系不同而不同。例如某位置是甲的家庭所在地，是乙的朋友的家，是丙偶尔路过的陌生地方，该位置对甲乙丙具有不同的社会属性。

2.1 自然属性

由于现在的各种感知设备基本都具有 GPS 定位功能，所以可根据 GPS 定位确定某个位置的自然属性，用 $A_{i,j}$ 来表示位置 l_j 对用户 i 的自然属性。

具有不同自然属性的地点，包含着人们不同的隐私信息，用户有不同的敏感程度。例如，医院、银行、公园、超市、野外、河流等这些地点的自然属性不同，医院对于病人来说，敏感程度非常高，病人通常不愿意泄露自己的健康状况；类似的，人们去银行办各种业务的相关隐私信息也是需要高度保密的；人们对超市、公园、野外或河流等这些地方的敏感度较低，因为这些地方包含会损害用户个人利益的隐私信息较少。因此，可以根据 GPS 定位信息初步判定某个位置的自然属性，再根据自然属性的现实意义给予一个自然属性敏感值 $A_{i,j}$ 。

2.2 社会属性

2.2.1 访问时长

通过对用户的移动轨迹进行分析，文献[16]发现每个用户每天的平均活动时间呈明显的幂律分布。根据用户在一段时间内的历史轨迹记录，统计出用户对每个位置的平均访问时长，可以反映出用户对该位置的依赖程度。若位置 A 是用户甲的家庭所在地，那么用户甲除了外出旅游或出差等特殊情况外，他

会经常出现在该位置。若位置A对用户乙没有特殊的社会意义，他只是偶尔路过这里，在一段时间内，他对该位置的访问时长会很短。由此可见，访问时长这一衡量指标可以在一定程度上反映出某位置对某用户的社会意义，从而判定该用户对该位置的敏感程度。

定义用户 i 对位置 l_j 的访问时长 D_{ij}：

$$D_{ij}=\frac{\int_{t_1}^{t_2} f(t)dt}{t_2-t_1} \tag{1}$$

$$f(t)=\begin{cases}1 & \text{在该区域}\\0 & \text{不在该区域}\end{cases} \tag{2}$$

其中：t_1 为访问记录的开始时间；t_2 为访问记录的结束时间。考虑到用户是不断移动的，并且这些位置的社会意义也有可能发生改变（比如用户搬家等），因此，随着用户移动轨迹的更新，用户的访问时长也可以实时动态地更新。

2.2.2 访问频率

用户 i 在（t_1，t_2）这段时间内的移动轨迹为 L_i =（$l_1,l_2,...,l_j,...,l_n$），$1\le j\le n$，$l_j\in L_i$，l_j 表示用户访问过的其中一个位置。文献[17]指出，根据访问频率在地点中的排名，可推导出位置的语义信息。若用户对某位置的访问频率较高，可以从一个方面反映出该位置对用户的重要性较高，从而判断该位置可能包含的用户隐私信息也较多。因此，本文将访问频率也作为判断用户敏感程度的一个衡量指标。随着用户移动轨迹的不断变化，对每个位置的访问频率也可以动态更新，达到实时预测用户对每个位置的敏感度的目的。

访问频率是指在一段时间内，用户对某个位置的访问频次占整个移动轨迹中对所有位置的总访问频次的比值。定义用户 i 对位置 l_j 的访问频率 $P(i,l_j)$ 为

$$P(i,l_j)=\frac{\mathrm{N}(l_j)}{\sum_{l\in L_i}\mathrm{N}(l)} \tag{3}$$

其中：$\mathrm{N}(l_j)$ 为用户 i 到达位置 l_j 的频次；$\mathrm{N}(l)$ 为用户 i 的整个移动轨迹中对所有位置的总访问频次。

2.2.3 访问的规律性

为了更准确地预测用户的敏感度，还要考虑另一个衡量指标——访问的规律性。访问的规律性反映用户对某位置的访问是否符合常态，从而排除偶然性的因素所导致的对用户敏感性的误判。可以考虑这样一种情况，某一感知用户只是偶尔在某个地方连续停留了几天，如旅游或出差，但其实他对该位置的敏感度并不高，却会计算出他对该位置的访问时长很高导致误判；或者他经常无规律地路过某个地方，会计算出他对该位置的访问频率很高，同样会导致误判。像家、工作地点等这些比较隐私的地方，用户通常是访问时长、访问频率都比较高，而且会很有规律地访问，而不是偶尔造访。

要计算用户对某位置访问的规律性，先算出用户与位置的平均分离周期，即他间隔多久会离开一次。

平均分离周期 AVG_{ij}：

$$AVG_{ij}=\frac{\int_{t_1}^{t_2}\delta_{i,j}(t)dt}{n_{i,j}} \tag{4}$$

其中：t_1 和 t_2 分别为某用户移动轨迹的开始时间和截止时间；$n_{i,j}$ 为用户 i 对位置 j 的分离次数。

$$\delta_{i,j}(t)=\begin{cases}0 & \text{用户}i\text{在}j\text{位置}\\1 & \text{否则}\end{cases} \tag{5}$$

本文采用高斯相似度函数对 AVG_{ij} 进行归一化，得到用户 i 与位置 j 的关系强度：

$$C_{i,j}=e^{-\frac{(AVG_{i,j})^2}{2\sigma^2}} \tag{6}$$

其中：σ 为分离周期的缩放参数。

为了能最终反映出用户对某个位置访问的规律性，测量分离周期的方差，使用不规则度量 $I_{i,j}$ 反映波动的大小（规律性）：

$$I_{i,j}=\frac{\sum_l(\mathrm{X}_l-C_{i,j})^2}{n_{i,j}} \tag{7}$$

其中：X_l 为分离周期的长度。

2.2.4 敏感等级函数

为了满足所有用户个性化的隐私安全需求，建立实时的、动态的隐私分级模型；同时考虑位置的自然属性、用户对某位置的访问时长、访问频率、访问的规律性这些衡量指标，定义了一个敏感等级函数，用它来判定用户对他访问的某个位置的敏感程度。对于某一感知任务，让对该位置敏感度较低的用户去参与感知；对某一感知用户来说，他对不同的位置有不同的敏感度，可以在敏感度较低的地方贡献自己的感知资源。因此，既可以保护用户的隐私信息，又可以在满足用户个性化的隐私安全需求的前提下实现感知资源的最大化利用，提供高精度的感知数据。

最终的敏感等级函数 $f_{i,j}$ 为

$$\begin{aligned}f_{i,j}&=\alpha P(i,l_j)+\beta D_{i,j}+\omega A_{i,j}+\mu I_{i,j}\\&=\alpha\frac{\mathrm{N}(l_j)}{\sum_{l\in L_i}\mathrm{N}(l)}+\beta\frac{\int_{t_1}^{t_2}f(t)dt}{t_2-t_1}+\omega A_{i,j}+\mu\frac{\sum_l(\mathrm{X}_l-C_{i,j})^2}{n_{i,j}}\end{aligned} \tag{8}$$

其中：α,β,ω,μ 为调节各个指标所占权重的参数。

2.2.5 用户-位置矩阵

根据用户对位置的敏感值建立矩阵 $\boldsymbol{M}$，矩阵的列代表多个位置，行代表多个用户，一个矩阵可以表示出一定范围内的用户与位置之间的敏感情况。对某一个位置来说，可以选择敏感等级较低的 h 个用户去完成感知任务；对某一个用户来说，可以选择敏感等级较低的 k 个位置去参与感知任务，贡献感知资源。矩阵示例如式（9）所示.

$$\boldsymbol{M}=\begin{matrix} & l_1 & l_2 & l_3 & \cdots & l_m\\ u_1 & f_{11} & f_{12} & f_{13} & \cdots & f_{1m}\\ u_2 & f_{21} & f_{22} & f_{23} & \cdots & f_{2m}\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ u_n & f_{n1} & f_{n2} & f_{n3} & \cdots & f_{nm}\end{matrix} \tag{9}$$

其中：f_{ij}（$1\leq i\leq n;1\leq j\leq m$）表示位置 l_j 对用户 u_i 的敏感等级值。

用户 u_i 在位置 l_j 对应的敏感阈值用 β_{nm} 表示，则用户一位

置对应的敏感阈值矩阵 β 可表示为

$$\beta = \begin{matrix} & l_1 & l_2 & l_3 & \cdots & l_m \\ u_1 \\ u_2 \\ \vdots \\ u_n \end{matrix} \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \beta_{23} & \cdots & \beta_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{n1} & \beta_{n2} & \beta_{n3} & \cdots & \beta_{nm} \end{bmatrix} \tag{10}$$

$$\beta_{ij} = 1 - f_{ij}; 0 \le \beta_{ij} \le 1 \tag{11}$$

2.3 实时动态的隐私分级算法

在上述各敏感参量统计计算方法的基础上，给出实时动态的隐私分级算法，以满足每个用户在不同位置的隐私安全需求，保证用户在不会泄露自己的隐私信息的前提下贡献感知数据；同时也在动态的、不可预测攻击手段的感知环境下通过实时的敏感度计算来为用户提供动态隐私判定方案，以保护用户的隐私，提高感知数据的精度。

实时动态的隐私分级算法如下：

a)根据感知设备上的 GPS 定位确定一些特殊位置的自然属性。

b)获得用户的历史轨迹记录，轨迹记录需要定期更新为最新记录，本文定为一周更新一次。

c)根据用户的访问记录统计出在这个周期内，用户访问每个地方的访问时长、访问频率、访问的规律性这些衡量指标，以表征某个位置对某个用户所具有的社会属性。

d)综合这个位置的自然属性和社会属性，根据敏感度函数计算出用户对该位置的敏感度。

e)在同一个位置，不同的用户有不同的敏感度，将用户与位置对应的敏感度值存入用户—位置矩阵，根据敏感等级得出用户—位置敏感阈值矩阵 β 。系统根据用户数量和感知任务的需求动态设定隐私阈值 θ（其中 $0 \le \theta \le 1$，$\theta = 0$ 表示无隐私保护，$\theta = 1$ 表示最大隐私保护）,若 $\beta_{ij} \le \theta$，则此用户在该位置可以参与感知任务而不会泄露隐私；否则该位置判定为用户的敏感位置，不能参与感知任务。

由于同一个位置对不同的用户来说有不同的现实意义，也就有不同的敏感度，可以让敏感度低的用户参与感知以保证感知任务能够完成；同一个用户在不同的位置有不同的敏感度，该用户可以在一些不会泄露个人隐私信息的位置贡献自己的感知资源，既保证不会泄露自己的隐私，又能实现感知资源的最大化利用。

3 实验结果与分析

3.1 实验数据的预处理

为方便描述，定义以下术语，包括 GPS 记录（P）、GPS 轨迹（$Traj$）、停留点（S）和位置历史（$LocH$）。

定义 3 GPS 记录。GPS 记录是 GPS 点 $P = \{P_1, P_2, \ldots, P_n\}$ 的集合。每一个 GPS 点 $p_i \in P$ 包含纬度（$p_i.Lat$），经度（$p_i.Lngt$）和时间戳（$p_i.T$）。

定义 4 GPS 轨迹。GPS 轨迹是根据其时间序列将这些 GPS 点连接成的曲线。如图 1 所示，如果两个连续的 GPS 点之间的时间间隔超过了一个确定的阈值 ΔT，就将这两个点分解在两条不同的 GPS 轨迹上。因此，$Traj = p_1 \to p_2 \to \ldots \to p_n$，其中 $p_i \in P$，$p_{i+1}.T > p_i.T$，$p_{i+1}.T - p_i.T < \Delta T$（$1 \le i < n$）。

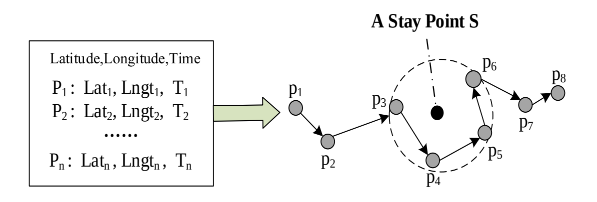


图 1　GPS 记录，GPS 轨迹和一个停留点

定义 5 停留点。停留点代表用户停留一段时间的地理区域，如图 1 所示的停留点 S。在本文中，每个停留点具有特定的语义含义，如生活或工作的地方、访问的餐厅和旅游的景点等。停留点的提取取决于两个参数，即时间阈值（$Tthreh$）和距离阈值（$Dthreh$）。与图 1 中所示的点$\{p_3, p_4, p_5, p_6\}$类似，单个停留点 S 可以认为是一组连续的 GPS 点 $P = \{p_m, p_{m+1}, \ldots, p_n\}$ 表示的虚拟位置，其中 $\forall\ m < i \le n$，距离 $D(p_m, p_i) \le Dthreh$，$|p_n.T - p_m.T| \ge Tthreh$。停留点 S 由 P、$Dthreh$、$Tthreh$ 三个因素决定。S 通常涉及到多个 GPS 点的空间区域，因此需要根据这些 GPS 点计算区域的平均坐标，$S = (Lat, Lngt, arvT, levT)$，即

$$s.Lat = \sum_{i=m}^{n} p_i.Lat / |P| \tag{12}$$

$$s.Lngt = \sum_{i=m}^{n} p_i.Lngt / |P| \tag{13}$$

$$s.arvT = p_m.T \tag{14}$$

$$s.levT = p_n.T \tag{15}$$

其中：$s.Lat$ 和 $s.Lngt$ 分别表示集合 P 的平均纬度和平均经度；$|P|$表示集合 P 中元素的个数；$s.arvT$ 和 $s.levT$ 分别表示一个用户在停留点 S 处的到达时间和离开时间。

定义 6 位置历史。位置历史是实体在一段时间内在地理空间中访问的位置的记录。在本文中，一个人的位置历史（$LocH$）表示他访问停留点的到达时间和离开时间相应的序列。

$$LocH = (s_1 \xrightarrow{\Delta t_1} s_2 \xrightarrow{\Delta t_2}, \ldots, \xrightarrow{\Delta t_{n-1}} s_n) \tag{16}$$

$$\Delta t_i = s_{i+1}.arvT - s_i.levT \tag{17}$$

其中：Δt_i 表示用户访问不同停留点之间的时间间隔。

3.2 实验环境

算法采用 C++实现，在 Intel(R) Core(TM) i3-2350M 2.3 GHz 处理器、4 GB 内存的 Windows 7 平台上运行。仿真数据集采用 GeoLife 项目[18]采集的真实数据集进行实验，该数据集中的数据点由不同采集频率的 GPS 记录器每间隔 2~5 s 采集一次，采集时间从 2007 年 4 月持续至 2012 年 8 月，数据集中共包含 182 个用户的 18 670 条 GPS 轨迹记录，共包含 2 487 万个数据

点，是典型的时空数据集。

停留点检测：在这个实验中检测停留点时，设置参数 $Tthreh=20\ minutes$，$Dthreh=200\ meters$。若一个用户在 200 m 范围内的区域停留时间超过 20 min，就将这个区域判定为一个停留点，即一个位置。

为了测试本文提出的 LPPA-PSRDU 算法的性能，在相同条件下，将本文算法与固定参数的静态位置隐私保护方案 Avg Static、Max Static 和自适应隐私保护策略的 Adaptive 算法进行仿真对比。分别从隐私保护水平、数据的完整性和数据的精确性等方面评估本文所提算法的有效性。各种算法的参数设置如表 1 所示。

表 1　算法参数设置

	Avg Static	Max Static	Adative	LPPA-PSRDU
θ	0.1-0.9	0.1-0.9	0.1-0.9	0.1-0.9
λ	1-10	1-10	Adaptive	
α				0.4
β				0.2
ω				0.2
μ				0.2

3.3　算法性能度量标准

本节中引入隐私保护水平度量标准来评估所提算法的有效性，通过度量感知数据的完整性和感知数据的精确性来比较此方案与之前方案的性能。

3.3.1 隐私保护水平

参与者的隐私泄露概率 $P_{Disclosure}$ 定义为位置隐私泄露的数量与需要保护隐私的位置总数量的比值。公式如下：

$$P_{Disclosure}=\frac{n_{Disclosure}}{N_{Disclosure}} \tag{18}$$

其中: $n_{Disclosure}$ 为被泄露的位置数量；$N_{Disclosure}$ 为需要保护隐私的位置总数量。隐私保护度 $Protection_Level$ 定义如下：

$$Protection_Level=1-P_{Disclosure} \tag{19}$$

3.3.2 数据的完整性

影响感知数据可用性的一个重要因素是感知数据的丢失。由于隐私问题，用户收集的一些感知数据可能不会上传给服务器。将数据的完整性 $Data_Completeness$ 定义为

$$Data_Completeness=\frac{\sum_{i=1}^{k}d_i(data)}{\sum_{i=1}^{N}D_i(data)} \tag{20}$$

其中：$\sum_{i=1}^{k}d_i(data)$ 为参与者提交给服务器的感知数据量；$\sum_{i=1}^{N}D_i(data)$ 为参与者感知到的总数据量。

3.3.3 数据的精确性

由于隐私保护问题，参与者提交不精确或粗粒度的位置信息，会导致感知数据的精度下降，从而影响数据的可用性。用感知数据对位置的平均绝对误差来衡量数据的精度（$Data_Accuracy$）。

$$Data_Accuracy=\frac{\sum_{i=1}^{N}\frac{|r_i'-r_i|}{r_i}}{N} \tag{21}$$

其中：r_i' 为提交的感知数据的位置；r_i 为感知数据的实际位置；N 为位置总数。

3.4　实验结果分析

算法的隐私保护水平如图 2 所示。随着隐私阈值的增大，四种算法的隐私保护水平均呈上升趋势。因为隐私阈值设置的越高，算法对参与者的位置隐私保护强度越大。但在相同的隐私阈值条件下，动态隐私保护机制 Adaptive 和 LPPA-PSRDU 明显比静态隐私保护机制 Max Static 和 Avg Static 的隐私保护水平要高。因为采用静态隐私保护策略时，默认所有参与者的所有位置均需要同等级别的隐私保护，采用预定义的固定静态参数，导致有些位置隐私保护不足，有些位置又保护过度。动态隐私保护策略可以由系统测算出实际需要的隐私保护等级，动态调整隐私保护参数，尽可能地满足多个位置不同的隐私保护需求。本文提出的 LPPA-PSRDU 算法比 Adaptive 算法的隐私保护水平更高，因为尽管 Adaptive 算法会动态测算参与者的隐私等级，但是其经过位置混淆处理后，依然会有一部分隐私泄露，而 LPPA-PSRDU 机制根据敏感等级动态的选择一部分敏感度低的用户参与感知任务，满足了用户不同的隐私需求。

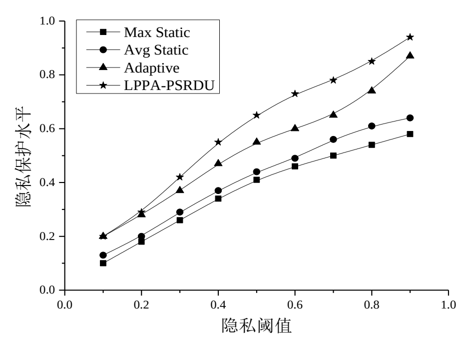


图 2　四种算法的隐私保护水平对比

感知数据的完整性如图 3 所示。当隐私阈值较低时，四种算法的感知数据完整性都较高；当隐私阈值较高时，数据丢失都较多。因为隐私阈值越高，需要隐藏的感知数据越多，以实现较高等级的隐私保护。总体上，两种动态的隐私保护算法在数据完整性上要优于两种静态的隐私保护算法。因为静态算法在整个感知过程中对所有位置都采用相同的隐私保护等级，导致参与者在一部分不敏感的位置被过度保护，浪费了感知资源。两种动态隐私保护算法中，LPPA-PSRDU 比 Adaptive 的数据完整性稍差一些，主要是因为 Adaptive 算法在隐私阈值高时，采用位置混淆机制上传感知数据，而 LPPA-PSRDU 算法只允许一部分符合要求的参与者贡献感知数据，另一部分敏感度过高的参与者在此位置不参与感知。

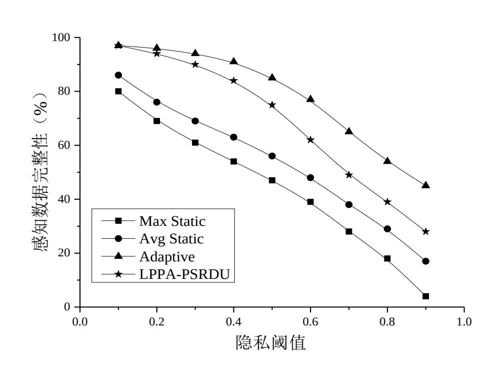


图 3 四种算法的感知数据完整性对比

四种算法的感知数据精确性如图 4 所示。随着隐私阈值的增大，Max Static、Avg Static 和 Adaptive 算法的数据精度都在逐渐下降，采用 LPPA-PSRDU 算法提交的感知数据精度一直趋近于 100%。这是因为其他三种算法都使用位置混淆机制，隐私阈值越大，该位置所需泛化的面积越大，导致数据的精度下降，而 LPPA-PSRDU 机制在满足参与者隐私需求的前提下，敏感度低的参与者可以提交准确的感知数据。LPPA-PSRDU 机制在时空相关性要求较高的感知场景下优势非常明显。

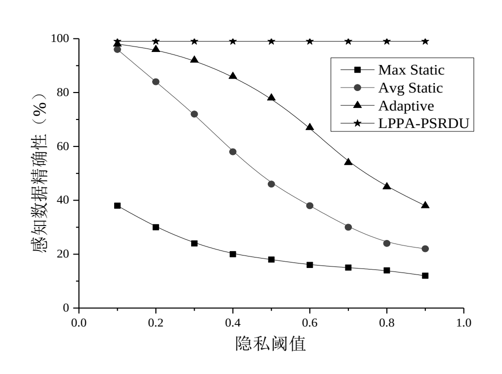


图 4 四种算法的感知数据精确性对比

4 结束语

针对群智感知网络中位置隐私泄露严重，现有算法对所有位置采用相同的隐私保护强度导致有些位置保护过度、有些位置保护不足，且获得的感知数据精度较低的问题，本文提出一种个性化的位置隐私保护算法（LPPA-PSRDU）。该算法根据用户的历史 GPS 轨迹数据计算出用户对不同位置的访问时长、访问频率和访问的规律性，挖掘出各个位置对于不同用户来说所具有的社会属性，再结合位置的自然属性，得到用户—位置敏感等级矩阵，在每个位置尽量让敏感等级低于隐私阈值的用户参与感知任务。该算法适用于对时空相关性要求较高的感知场景，在满足用户在不同位置的隐私安全需求的同时，能获得高精度的感知数据。通过与两种静态位置隐私保护算法（Avg Static 和 Max Static）和一种自适应位置隐私保护算法（Adaptive）对比，验证了本文所提算法在隐私保护水平、数据完整性和数据精确度方面的性能均优于 Avg Static 和 Max Static；本文算法与 Adaptive 算法相比，损失了一部分数据完整性，但是获得了较高的隐私保护水平和数据精确度。在参与感知的用户较多、对感知数据的时空相关性要求较高的场景下，本文提出的算法是非常有意义的。当然，该方法也存在一些不足，例如在评估用户对某位置的敏感性时，需要较大的计算量。如何快速准确地评估出用户的敏感性，较好地保护用户需要保护的隐私，是未来的工作重点。

参考文献：

[1] 吴垚，曾菊儒，彭辉，等. 群智感知激励机制研究综述 [J]. 软件学报, 2016, 27 (8): 2025-2047.

[2] Sun W, Liu J. Congestion-aware communication paradigm for sustainable dense mobile crowd-sensing [J]. IEEE Communications Magazine, 2017, 55 (3): 62-67.

[3] 张学军，桂小林，伍忠东. 位置服务隐私保护研究综述 [J]. 软件学报, 2015, 26 (9): 2373-2395.

[4] Shen H, Bai G, Yang M, et al. Protecting trajectory privacy: a user-centric analysis [J]. Journal of Network & Computer Applications, 2017, 82: 128-139.

[5] Mehta K, Liu D, Wright M. Protecting location privacy in sensor networks against a global eavesdropper [J]. IEEE Trans on Mobile Computing, 2011, 11 (2): 320-336.

[6] Yang D, Fang X, Xue G. Truthful incentive mechanisms for k-anonymity location privacy [C]// Proc of INFOCOM. Turin: IEEE Infocom, 2013: 2994-3002.

[7] Ziegeldorf J H, Henze M, Bavendiek J, et al. TraceMixer: privacy-preserving crowd-sensing sans trusted third party [C]// Proc of Wireless On-demand Network Systems and Services. Jackson, WY: IEEE Press, 2017: 17-24.

[8] Yu Z T, Qian Q, Lin C Y, et al. High performance datafly based anonymity algorithm and its l-diversity [J]. International Journal of Grid and High Performance Computing, 2015, 7 (3): 85-100.

[9] 张伊璇，何泾沙，赵斌，等. 一个基于博弈理论的隐私保护模型 [J]. 计算机学报, 2016, 39 (3): 615-627.

[10] Zhang Ning, Li Ming, Lou Wenjing. Distributed data mining with differential privacy [C]// Proc of IEEE International Conference on Communications. Kyoto: IEEE Press, 2011: 1-5.

[11] Li M, Zhu L, Zhang Z, et al. Achieving differential privacy of trajectory data publishing in Participatory Sensing [J]. Information Sciences, 2017, 400-401: 1-13.

[12] Wei Z, Zhao B, Liu Y, et al. PPSense: a novel privacy-preserving system in people-centric sensing networks [C]// Proc of the 8th International ICST Conference on Communications and Networking in China. 2013: 461-467.

[13] Liu Bozhong, Chen Ling, Zhu Xingquan, et al. Protecting location privacy in spatial crowdsourcing using encrypted data [C]// Proc of the 20th International Conference on Extending Database Technology. Venice, Italy:

OpenProceedings, 2017: 21-24.

[14] Krumm J. A survey of computational location privacy [J]. Personal and Ubiquitous Computing, 2009, 13 (6): 391-399.

[15] Agir B, Papaioannou T G, Narendula R, et al. User-side adaptive protection of location privacy in participatory sensing [J]. Geoinformatica, 2014, 18 (1): 165-191.

[16] Henderson T, Kotz D, Abyzov I. The changing usage of a mature campus-wide wireless network [J]. Computer Networks, 2008, 52 (14): 2690-2712.

[17] Prabhala B, Porta T L. Spatial and temporal considerations in next place predictions [J]. Proceedings IEEE INFOCOM, 2015, 2015 (9): 390-395.

[18] Zheng Y, Xie X, Ma W Y. GeoLife: a collaborative social networking service among user, location and trajectory [J]. Bulletin of the Technical Committee on Data Engineering, 2010, 33 (2): 32-39.